



June 2004

Taxonomy Skills for Search Teams

This article originated as a member question: should the teams responsible for the search function work with people who have taxonomy/metadata skills? To make the argument, she wanted to find out how other companies organize and staff their intranet/Web portal search and taxonomy functions. To help her, we sent an e-mail asking our contacts the following questions:

In your company, are different sets of people involved in the search and taxonomy functions? If so, how well do they collaborate regarding requirements, deliverables, goals, priorities, etc.?

In Part I, we summarize the responses, most of which support the requestor's position. Two people indicated that technology selection played a primary role in getting search and taxonomy specialists to collaborate. Another noted the importance of a liaison person with both taxonomy and IT background. However, even the most enthusiastic proponents of collaboration lamented the need to constantly sell taxonomy benefits to users and business units (see also "Selling taxonomy: a blind alley?").

A more interesting issue is why this question evoked such an extensive response — and why it needs to be asked in the first place. We think this is because information management in large companies is often technology-focused rather than people-focused. Most employees understand the value of A - Z indexes, directories, tables of contents, searches, and spreadsheets. If your starting point is user productivity and you avoid techie language, are concepts like metadata repositories, taxonomies, and fielded searches really that hard to appreciate?

In Part II, we look for clues about why the issue seems to be a problem for so many people. Three possibilities emerge:

- search and taxonomy teams have different mandates and different short term objectives;
- the connection between the tasks and the bottom line are remote;
- the editorial function, which normally serves as an advocate for readers, is weak.

PART I: RESPONSES

1. *Web site coordinator, financial services company*

Duh! It makes so much sense it must hurt to think about it any other way.

Our Search/Taxonomy Team has access to the following staff:

- 4 librarians with Masters of Library Science degrees and various subject specialties;
- our "search platform" administrator, who maintains the servers, connections, processing, and the software;
- a content/metrics expert;
- a user interface designer (a recent addition).

These, of course, are not full time positions, but when we convened to select a new tool these are the critical support personnel involved in making the decision. To pick an upgraded search platform, we also supplemented the Search/Taxonomy team with:

- integration specialists;
- metadata specialists;
- enterprise content management specialists;
- enterprise architecture and security specialists.

We will be developing a system blueprint to document how search integrates with various applications and web implementations as well as the taxonomy maintenance modules and indexes required.

When we did our search optimization project a few years ago, we spent most of our time working on content metadata to make it work

better with the search engine, not the other way around. Here is the metaphor that I use. It may be better understood in Northern climates: Content is like a patch of ice on a road, the search engine is like the tire of a car. Metadata (of which taxonomy is a component) is the salt that gives the tire traction.

Now, this enviable position of working with the search platform support team was not easily accomplished. It took major dissatisfaction with the existing search function to make upper management listen. All of the consulting groups such as Forrester support a combined search/taxonomy team. In fact, the new tools demand this collaboration. Search is surprisingly complex these days, closely integrated into corporate content and not just a technical implementation. It is very user-centric, so end user involvement and testing is important.

We use standard project management protocols for projects. Charters, project plans, milestones, deliverables with funding. We also have support budgets for daily feeding and maintenance. We use our intranet to publish team related information and collaboration.

One additional twist is that even if the Search and Taxonomy people collaborate, how do you get your user groups to engage this combined team?

2. Consultant, energy company

For existing web pages and portals, the Java programmers and a lead coordinator work with the infrastructure departments such as Legal, HR, Trading, Environmental, Risk, and Engineering to develop an index of topics. Once the topical index is determined, topics are linked to documents. There are no specialized taxonomists involved, and the first draft usually requires reengineering because of oversights.

The developers use a collaboration tool called eRoom (now owned by Documentum). It is a web-based tool that provides locations (eRooms) for folders and index databases. Collaborating users store shared documents and email. They also may receive shared

work assignments and track changes to project status.

In a current consulting assignment, I'm working with a company that is attempting to meet the requirements of the Sarbanes-Oxley law. The law requires that organizations establish and document internal controls along with the policies and procedures to support them. Teams have been assigned in all of the departments to develop controls and to determine gaps and improvements.

To see that all bases are covered, the Internal Controls director and CFO must be able to track the implementation and testing of the controls. The logical solution is a taxonomy and database for project control. This would enable the company to index and share documentation with the internal and external auditors as well as the infrastructure teams.

3. Plain language specialist, financial services company

I agree with the Society member's view that search function teams should work with people who have taxonomy and metadata skills. My remarks are based on my experience helping to develop a new corporate intranet for two merging companies.

Previously each company had its own intranet. The smaller company actually had more material on its intranet (procedures, policies, forms, general company information). The taxonomy was based on a community metaphor to encourage communities of practice just starting in the company. It had high level categories called City Centre (finance), Library (reference), Knowledge Depot (learning) etc. and wasn't very successful. People who didn't appreciate the metaphor or found searching frustrating, were less than thrilled.

The larger company had mostly communications on its intranet and a taxonomy invented by their communications department. It had four main categories - need to know, nice to know, did you know, and let us know (a weekly quiz). This taxonomy worked reasonably well for communications (once people figured out which news

was under which heading) but was useless for other types of information.

Neither of these intranets, to my knowledge, had the benefit of people with taxonomy or metadata skills.

The new combined intranet had two teams - a team with taxonomy and metadata skills and a technical team. The taxonomy team had three skill sets:

- a person with plain language and taxonomy experience (me);
- a person who had both programming and taxonomy knowledge;
- an intranet manager who was the gatekeeper for intranet categories, headings, etc.

These three people were responsible for developing the draft taxonomy, refining it according to business unit needs, helping business units decide what to move to the new intranet and where to place it, and evaluating requests for new categories.

The programming team was concerned with IT issues, and the search function. I don't know as much about this area.

The two teams collaborated quite well, mostly because the person on the taxonomy team who had technical skills acted as a liaison. To save time, it's important to have a central decision-clearing process, but the main job of the taxonomy team was to remain true to the "vision" of a corporate intranet and its long-term viability. The technical team wanted to "get going" and get things done. I don't think this difference in focus is that unusual, but it's good to be aware of it and to have buy-in for the vision of an intranet at a high level in the corporation.

Constantly, we found it was important to "sell" the benefits of a taxonomy, to explain what it was, and to control where intranet content would live (in terms of categories, not physical storage). We couldn't be dogmatic about this, but we often had to be tough with business areas, in order to maintain the overall integrity of the categories. Everyone thinks their content is the most important ...

Regarding requirements and deliverables, the taxonomy team wrote a requirements/scoping paper early in the process and got buy-in from all the

stakeholders. This is the point at which expectations have to mesh. After all, a taxonomy team is nothing without the technical team to make it happen. Conversely, a technical team without a taxonomy team can design an “intra-mess” which ends up frustrating everyone.

I know that there were differences in priorities. One difference was that the parent company wanted high level categories named for the various business units. The taxonomy team, on the other hand, wanted to organize by topic. In the end, the search function was able to accommodate both ways.

Differences may have been exacerbated somewhat because two cultures were merging, and it was the smaller company who had actually done more work on taxonomy. This sometimes led to situations of “Who bought whom? Who’s telling whom what to do?” But ultimately, most decisions were made democratically and with any eye to having the “best” intranet possible.

The new intranet has been in place for 6-8 months now, and to my knowledge things are working well.

4. Information architect, government agency

In the State of Minnesota, we have eight theme managers that keep the taxonomy current. Since our taxonomy is driven by Ultraseek’s CCE (Content Classification Engine), it is necessary to coordinate the search functionality with taxonomic functions. That is part of my role as the CCE administrator and overall taxonomy expert. For details, see “Integrating taxonomies in the state of Minnesota.”

5. Chief knowledge officer, financial services company

Here are a few quick answers to your questions.

a. How do we develop a common conceptual framework? Very difficult, usually just the word taxonomy baffles people.

b. How do we develop a common vocabulary? Documentation specialists edit and police the terminologies of others. As organizations grow in size

and across regions, this is an increasing challenge.

c. How do we define our domains? Use common sense applied to experience with a sense of what the overall goal of the taxonomy is.

d. How do we distinguish between a taxonomy and a taxonomy architecture? Once users see results of a taxonomy approach popping up in different places this is usually not a problem.

e. What tools and techniques are available for linking specialized taxonomies? We use databases and web pages.

f. What role does content creation play in taxonomy maintenance, and how can the system encourage authors to use taxonomy data? Usually tagging of content with codes has to be done by specialists. Normal users don’t usually understand the implications of selecting the codes, unfortunately.

g. What methods are available to extract, exchange, and use taxonomy data stored in proprietary data formats? Somehow you need to store metadata. We don’t try to add metadata within a document.

h. What are the pros and cons of metadata repositories and solutions? I don’t see any cons. The basic pro is to link content from different taxonomies.

i. How should we design, implement, and analyze a taxonomy pilot? Just do it and let the benefits display themselves so that usage can gradually spread.

j. How do we assess our organization’s readiness for taxonomy development, and what are the options for a phased implementation? If there is a need to standardize and distribute information for different activities using IT tools, you could probably benefit from some form of taxonomy.

You could argue that simply storing files in folders with names is a form of taxonomy. So “implementing taxonomy” could be perhaps replaced with “developing your taxonomy.” Perhaps a major step is considering different dimensions — i.e. a folder structure is typically a one dimensional solution. Tagging info with standard codes introduces a new dimension. Discipline in few dimensions is definitely better

than laxity in many.

6. Taxonomist, pharmaceutical company

Our taxonomy and search engine software both are supported by Convera. The search engine is Retrievalware 8.X, and the taxonomy is supported by the Dynamic Categorization and Classification tool. We have the same technical team responsible for upgrades to the search engine and the taxonomy data.

Our search engine and taxonomy tools each have different system owners; both are members of the other’s team so there are no communication problems. Together, we set the goals and priorities for improving the behavior of the search engine as well as working on the taxonomy project.

The system owners right now are the project leaders for the systems; we are both in the Information Management department. We have different projects in place to improve the search engine which are separate from any projects involving the taxonomy team.

As the taxonomy owner, I have been involved with the search engine since we were choosing the tool. Similarly, the search engine owner was involved with the search for the taxonomy tool which was conducted a few years later. We knew that the two tools had to work closely together, and we preferred that the taxonomy tool use the index generated by the search engine. I had led a team earlier which created a “directory of information” for the intranet so I was chosen to lead the taxonomy teams.

The search engine owner and I work together. Having two different people responsible for the search and taxonomy allows us to run more projects.

The search index isn’t the basis for the taxonomy. Unless the taxonomy indexer and the search engine indexer are the same (or can read each other’s data), all of the data would need to be indexed twice (once for taxonomy and once for search). This means that the two might not be synchronized. When we were looking at tools, all of the taxonomy companies who also

had a search engine were suggesting that either we use their search engine, worry about synchronization, or go through a painful process to get the two indices synchronized. We had already purchased Convera's Retrievalware as the search engine and this was in place to be used throughout the corporation for many data repositories. We did not want to purchase another search engine just to have taxonomy. About 18 months after we began our search, Convera released their taxonomy solution. We had the money, it fit most of our requirements, so we bought it.

We have a 'data dictionary', which Convera calls a 'Domain Cartridge' or a 'User-defined cartridge', that allows us to create related concepts. Any taxonomy we create is stored in a 'taxonomy cartridge'. During indexing, key concepts, relationships and linkages are extracted based upon taxonomic structures. With this we can retrieve relevant search results even when the search terms may be different from, but related to, the original source.

When it comes right down to it, the search index itself is just a list of words which connects with the documents associated with them. When a user performs a search, the query is reformulated to retrieve all relevant documents based on the data dictionary. RetrievalWare's Dynamic Classification engine organizes search results on-the-fly into selectable classification views. These classification views are based upon taxonomies. The graphic below illustrates the process.

7. Taxonomist, software company

The answer is YES!! To me this is a no-brainer. But time and time again I find myself educating and reminding people that the entire system needs to be considered. Search does not live by itself. This is a soapbox for me. You hit a hot button.

I work as a taxonomist on the User Assistance team, which creates the help systems in software products. The UA team basically consists of writers, editors, taxonomist (me), search author and data analyst. I am two doors down from the search author, and work with

this person daily.

Besides my immediate team I work cross-team with usability researchers and designers. Together we are helping to define the user experience for "finding" information in the help system, specifying requirements for tools to create the information architecture needed, and then creating the information architecture.

Findability is about all methods of finding - search, browse, and asking. The individuals working on the different areas need to be working together to create the desired user experience.

In addition, refer to the 2nd edition of the Morville and Rosenfeld book, Information Architecture for the World Wide Web. Their recommendation is a mixed skill set for any team creating a web-based information system.

This question reminds me of the decades old battle in the traditional librarian space of "should reference librarians understand cataloging and should catalogers understand interfacing with the customer as a reference librarian." YES!!

PART II: OUR COMMENTS

What is most interesting about these comments is why the question needs to be asked in the first place? Why is the issue of search/taxonomy collaboration a "hot button" for one respondent, and why does another conclude her remarks by asking "How do you get your user groups to engage this combined team?"

We get one clue from respondent #3, who says, "The main job of the taxonomy team was to remain true to the 'vision' of a corporate intranet and its long-term viability. The technical team wanted to "get going" and "get things done." Different teams have different objectives and consequently different perspectives.

But that can't be the whole story. We get another clue from respondents #2 and #7, who mention specific applications. Why can't respondent #2 show the CFO a demo illustrating not only how taxonomy-enriched search can make Sarbanes-Oxley compliance easier but also how it can lower com-

pliance costs? Why can't respondent #7 show the help system editor how collaboration between the search and taxonomy teams can build customer loyalty and reduce customer support costs? In other words, the connection between the teams' actual work and the bottom line benefit is either nonexistent or too remote.

We get a third clue from respondent #5, who says, "the word 'taxonomy' baffles people." Of course it does, but everyone understands the value of a table of contents, a back-of-the-book index, and a Google search. With a little patience, it's not too hard to show a specific example illustrating the value — and the limitations — of each tool. The idea of combining the virtues of all three tools is easy for even the most technically naive person to grasp. Or, as respondent #1 says, "Duh! It makes so much sense it must hurt to think about it any other way."

Where are the editors?

The final clue comes from respondent #7, the only one to mention writers and editors as part of the team. Editors can fill two key roles:

1. represent the interests of the reader;
2. help the publisher make cost/benefit decisions.

We think that editors are under-represented in the search/taxonomy debate for 3 reasons:

1. Professionally trained editors (e.g. those in corporate communications) lean toward a print-based publishing paradigm that gives too much emphasis to a single publication and not enough to document collections, access to experts, and social networks.

2. De facto departmental editors lack professional status and training, even though they may be very proficient in certain publishing tasks. To compound the problem, many department and business unit managers do not view themselves as publishers, a key role in creating intellectual assets for the firm.

3. Most companies lack an Editor-in-Chief role, which serves as an advocate for enterprise-wide audiences and provides both direction and support

for departmental editors.

Our solution to the search/taxonomy divide is neither radical nor expensive. Find a bottom line issue that either saves money or makes money. Develop a prototype with an A - Z index, table of contents, and taxonomy-enhanced search. Develop a short demo and presentation for the person who “owns” the issue. Implement the prototype as an operational system for that issue, then generalize the system for other issues.

UPDATE (as of November, 2006)

Here’s what the IT Team Lead in a major computer industry company had to say about his organization’s search staff:

“For the customer search experience on our public Web site, we have a Business Lead, an IT Manager (who does manager things for the IT team), an IT Team Lead (technology expert), and IT team members who do support, development, and search engine administration.

The Business Lead does Search Engine Optimization stuff. He has business marketing skills (external company search is mostly about marketing from a business perspective), web marketing skills (i.e usability, click thru, abandonment, etc.) This person understands our company’s web strategy and how search can be used to promote/strengthen it. He identifies key business leads for all the major publishers in the company (i.e. store sites, support sites, etc.) and understands the behavior of search users, and how different search features are (or are not) useful for certain audiences.

The IT Team Lead is a technical expert on Search Technologies and general IT technologies (e.g. web services, system architecture, etc.). He handles delivery of search IT projects, including all the back-end integrations. He must remain aware of changes in the enterprise search space from a technology perspective.

Both the Business Lead and the IT team Lead are fulltime employees. Neither has much formal authority, but both have a lot of influence on what the search experience will be and what search technologies will be used.”

The Montague Institute Review is published by the Montague Institute and edited by Jean Graef.

© Copyright 1998 - 2015 Jean L. Graef. All rights reserved.